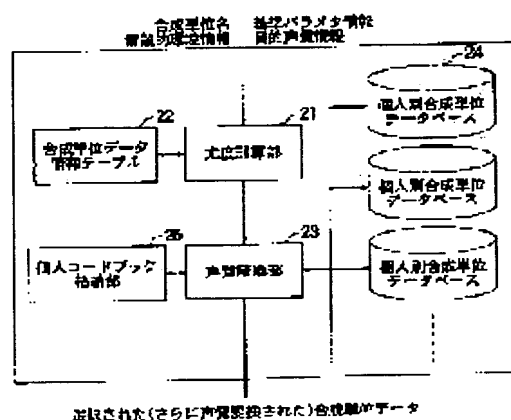


(11)Publication number : 07-319495
(43)Date of publication of application : 08.12.1995

G10L 3/02
G06F 17/21

(72)Inventor : SHIROTSUKA OTOYA
MURAKAMI NORIYA
HAYASHI KEIJI



THIS PAGE BLANK (USPTO)

of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

THIS PAGE BLANK (USPTO)

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平7-319495

(43)公開日 平成7年(1995)12月8日

| | | | | |
|-------------------------|------|---------|----------------|---------|
| (51)IntCl. ⁶ | 識別記号 | 庁内整理番号 | F I | 技術表示箇所 |
| G 1 0 L 3/02 | A | | | |
| G 0 6 F 17/21 | | 9288-5L | G 0 6 F 15/ 20 | 5 6 8 Z |

審査請求 未請求 請求項の数 2 F D (全 5 頁)

(21)出願番号 特願平6-136401

(22)出願日 平成6年(1994)5月26日

(71)出願人 000102728

エヌ・ティ・ティ・データ通信株式会社
東京都江東区豊洲三丁目3番3号

(72)発明者 城塚 音也

東京都江東区豊洲三丁目3番3号 エヌ・
ティ・ティ・データ通信株式会社内

(72)発明者 村上 憲也

東京都江東区豊洲三丁目3番3号 エヌ・
ティ・ティ・データ通信株式会社内

(72)発明者 林 慶士

東京都江東区豊洲三丁目3番3号 エヌ・
ティ・ティ・データ通信株式会社内

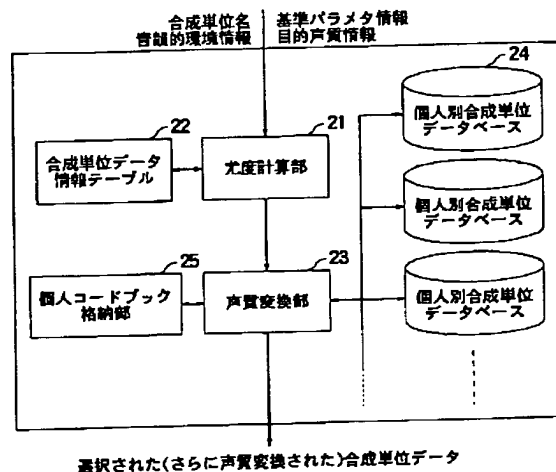
(74)代理人 弁理士 上村 輝之

(54)【発明の名称】 音声合成装置のための合成単位データ生成方式及び方法

(57)【要約】

【目的】 音声合成装置において、声質の異なる複数の人間の音声データを用いて、所望の声質を持った合成音声を得られるようにする。

【構成】 複数の人間の音声から得た合成単位データをデータベース25として蓄積しておく。音声合成したいテキストと声質が入力されると、そのテキストを構成する各音声合成単位に対して、音韻的環境、パワー、ピッチ及び声質に関し、総合的に最も近い合成単位データをデータベース25から選択する。次に、選択した合成単位データの話者の声質と、合成したい声質とがある程度以上異なるか否かチェックし、異なる場合には、合成単位データに声質変換を行って、合成したい声質と同様な声質に変換する。



【特許請求の範囲】

【請求項1】 複数話者の音声から得た合成単位データを蓄えるための合成単位データ保持部と、音声合成に用いたい合成単位と声質とを指定した情報を入力し、指定された合成単位に最も近い合成単位データを、前記合成単位データ保持部から選択する合成単位データ選択部と、

前記選択した合成単位データの話者の声質が前記指定された声質に対し所定程度以上異なるかどうかチェックし、異なる場合、前記選択した合成単位データの声質を前記指定された声質に近づけるための声質変換を行う声質変換部と、を備えることを特徴とする音声合成装置のための合成単位データ生成方式。

【請求項2】 音声合成に用いたい合成単位と声質とを指定する過程と、予め蓄積された複数話者の音声から得た合成単位データの中から、前記指定された合成単位に対し一致度の最も高い合成単位データを選択する過程と、

前記選択した合成単位データの話者の声質が前記指定された声質に対し所定程度以上異なるかどうかチェックし、異なる場合、前記選択した合成単位データの声質を前記指定された声質に近づけるための声質変換を行う過程と、を備えることを特徴とする音声合成装置のための合成単位データ生成方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、一般には、作成した文章の校正や視覚障害者の読書等に活用される、電子化されたテキストを自動的に読上げるための音声合成装置に関し、特に、合成単位データの生成技術の改良に関する。

【0002】

【従来の技術】従来の音声合成装置では、品質の良い合成音声出力するために、音声の合成単位に、その合成単位に前後の音韻的環境を考慮したサブカテゴリを設ける必要があった。たとえば「朝日」/asahi/は合成単位で記述すると/a//sa//hi/となるが、その合成単位/asahi/の/sa/と同様の先行音韻/a/と後続音韻/hi/に挟まれて使用された音韻データを必要とする。たとえば「旭川」/asahikawa/の/sa/は/asahi/の/sa/と同様の音韻的環境にあり、単語/asahi/の合成に使用することができる。しかし、合成単位のすべてに対して多数のサブカテゴリを設けるために、音声合成単位の作成に必要な音声データが大量に必要であった。また、合成する単語に求められる、音声のピッチ、パワーといったパラメータと、合成単位データのもつパラメータに違いがある場合、合成音声の品質が著しく悪化する。そのために同じ音韻的環境をもつ合成単位であっても、パワー、ピッチの異なるものを数種類もつ必要があり、合成単位作成のためのデータが大量に必要であった。

【0003】一方、音声データの声質を変換する技術が報告されている。たとえば、阿部伸「音声変換技術」信学技報Vol. 93, No. 427, pp69~pp75では、音声の音原特性と声道特性のそれぞれに注目した声質変換方法が紹介されている。

【0004】

【発明が解決しようとする課題】音声合成装置では、前項で述べた理由により大量の音声データを合成単位の作成のために必要とする。そこで、種々の場面で収録された複数の人間の音声データを使用して合成単位を作成し、それをつなぎあわせ音声合成をおこなうことが考えられるが、そうすると、個人ごとに音声の声質が異なるために、合成音声の品質が悪化する。そのため、合成単位作成のための音声データは、ある1人の人間から長時間にわたって収録する必要がある、収録作業に多大な労力が必要であった。

【0005】従って、本発明の目的は複数の人間の音声から合成単位に使用する音声データを選び、それをつなぎあわせて音声合成をしても、1人の人間の音声から音声合成した場合に近い合成音声を得られるようにすることにある。

【0006】

【課題を解決するための手段】本発明の合成単位データ生成方式は、複数話者の音声から得た合成単位データを蓄えるための合成単位データ保持部と、音声合成に用いたい合成単位と声質とを指定した情報を入力し、指定された合成単位に最も近い合成単位データを、合成単位データ保持部から選択する合成単位データ選択部と、選択した合成単位データの話者の声質が指定された声質に対し所定程度以上異なるかどうかチェックし、異なる場合、選択した合成単位データの声質を指定された声質に近づけるための声質変換を行う声質変換部とを備える。

【0007】

【作用】複数話者の音声から得た合成単位データがデータベースに予め蓄積されている。音声合成に用いたい合成単位と声質とが指定されると、まず、指定された合成単位に最も近い合成単位データが、データベースから選択される。次に、その選択した合成単位データの話者の声質が指定された声質とどの程度異なるかチェックされ、所定の程度以上異なる場合には、指定声質に近くなるように、合成単位データに対し声質変換が行われる。このようにして、声質の異なる種々の話者の音声データから、所望の単一の声質の合成単位データを得ることができる。

【0008】

【実施例】以下、本発明の実施例を図面により詳細に説明する。

【0009】図1は本発明に係る音質変換型音声合成装置の一実施例の全体構成を示す。

【0010】まず、装置の構成を説明する。この音質変

換型音声合成装置は入力端子11より漢字かな混じり文を入力すると共に、合成したい声質（性別、年齢層、声の高低等の指定であり、以下、目標声質という）を指定する。前処理部12は、この入力文を言語的に解析し、文を構成する単語の音韻記号系列と、単語のアクセント型と、入力文の文型とを求める。選択基準パラメタ設定部13は、求められた単語の音韻記号系列、単語のアクセント型及び入力文の文型から、音声の合成に必要な合成単位（以下、対象合成単位という）とその音韻的環境（＝先行音韻、後続音韻）と所定のパラメタ（＝パワー、ピッチ）を決定する。

【0011】合成単位データ選択部14は、予め合成単位ごとに複数の人間の音声より得られた複数の合成単位データを保持し、選択基準パラメタ設定部13から渡される対象合成単位の音韻的環境及びパラメタと目標声質とに最もよく合致する合成単位データを選び出す。更に、その選択した合成単位データの声質が目標声質とどの程度異なるかをチェックし、ある程度以上異なる場合、その合成単位データを目標声質と同様の声質に変換する。

【0012】合成単位データ変形部15は、選択された合成単位データを、対象合成単位のパラメタに合うように変形する。合成単位データ接合部16は、変形された合成単位データを接合して単語データを合成する。出力端子17は、合成された単語データを音声に変換し出力する。

【0013】次に、この音声合成装置の特徴である合成単位データ選択部14の構成を、図2を参照して詳しく説明する。

【0014】図2において、尤度計算部21は、選択基準パラメタ設定部13より渡される対象合成単位の合成単位名、音韻的環境、パラメタ及び目標声質に基づいて、これらの点で対象合成単位に最もよく一致する合成単位データの識別番号（以下、データIDという）を、合成単位データ情報テーブル22及び個人コードブック格納部25を参照して、決定する。

【0015】ここで、合成単位データ情報テーブル22には、図3に示すように、システム内に格納されているすべての合成単位データについて、その合成単位名、音韻的環境（先行音韻、後続音韻）、パラメタ（パワー、ピッチ）、話者識別番号（以下、話者IDという）、及びデータIDが記録されている。また、個人コードブック格納部25には、システム内に格納されている合成単位データのすべての話者に関して、話者IDとその声質の特徴を表わした情報とを記録した話者コードブックが格納されている。

【0016】尤度計算部21の処理を具体的に説明すると、まず、合成単位データ情報テーブル22から各合成単位データの音韻的環境とパラメタとを取り出し、また、個人コードブック格納部25から各合成単位データ

の話者の声質に関する情報を取り出す。次に、選択基準パラメタ設定部13から渡された対象合成単位の音韻的環境、パラメタ及び目標声質情報と、先程合成単位データ情報テーブル22及び個人コードブック格納部25から取り出した各合成単位データの音韻的環境、パラメタ及び声質情報とを、以下に示す評価関数に代入して、対象合成単位と各合成単位データとの間の距離Dを求める。

$$D = W1 \cdot Pre(s1, t1) + W2 \cdot Post(s2, t2) + W3 \cdot Pow(s3, t3) + W4 \cdot Pit(s4, t4) + W5 \cdot Person(s5, t5)$$

ここに、s1～s5は、対象合成単位の音韻的環境、パラメタ及び目標声質情報、t1～t5は、合成単位データの音韻的環境、パラメタ及び声質情報、Pre()は、先行音韻に関する距離計算関数、Post()は、後続音韻に関する距離計算関数、Pow()は、パワーに関する距離計算関数、Pit()は、ピッチに関する距離計算関数、Person()は、声質情報に関する距離計算関数、W1～W5は、重み係数である。この評価関数は、先行音韻、後続音韻、パワー、ピッチ及び声質の各々について、対象合成単位と合成単位データとの間の距離を計算し、それぞれの距離に、合成音声の品質に対する影響度を考慮した重み係数を掛け、それらの総和をとったものである。

【0018】尤度計算部21は、この距離Dの計算を、システムに格納されているすべての合成単位データについて行い、最も距離Dの小さい、つまり一一致度（尤度）の高い合成単位データを音声合成に使用するものとして決定し、その合成単位データのデータIDと話者IDと上記計算の過程で求めた声質に関する距離（一致度）とを、声質変換部23に渡す。

【0019】声質変換部23は、尤度計算部21から渡されたデータIDと話者IDに基づき、これにより特定される実際の合成単位データを個人別合成単位データベース24から取り出す。ここで、個人別合成単位データベース24には、各話者毎にその話者の収録音声から得た合成単位データが蓄積されている。続いて、声質変換部23は、尤度計算部21から渡された声質の一致度について、所定の閾値より低いか否かチェックし、低い場合（声質が相当に異なる場合）には、指定声質とほぼ同様の声質になるように、取り出した合成単位データの声質変換を行う。

【0020】この声質変換の手法としては、例えば、前掲「音声変換技術」に記載の声道特徴を考慮した話者コードブックを用いた方法を採用する。即ち、個人コードブック格納部25から、上記話者IDに対応する話者のコードブック（以下、オリジナルコードブックという）と、目標声質情報にマッチする声質を持った話者のコードブック（以下、目標コードブックという）とを取り出し、オリジナルコードブックから目標コードブックへコードブックマッピングを行うことにより、合成単位データの声質を目標声質に変換する。

【0021】尚、声質変換の手法には他にも種々のものが知られており、それを採用しても構わない。例えば、粕屋、楊著「音原特性を考慮した声質変換」(音響学会講論集、1-6-18, pp225-226, 1991)に記載の方法等が採用できる。

【0022】以上の処理により、複数の話者から収録した音声データを利用して所望の声質を持った音声合成可能になる。そのため、一人の人間から大量の音声データを収録する必要がなくなり、すでに収録されている種々の人間の音声データを寄せ集め再利用することができる。

【0023】本発明は、上述した実施例以外にも種々の態様で実施することが可能である。例えば、上述した評価関数において、声質に関する一致度のスコアに対する重み係数 $W5$ を適度に大きくすることにより、声質の大きく異なる話者の合成単位データが選ばれなくすることができる。これにより、声質変換部23を省略することも可能である。

【0024】また、合成単位データ選択部14は、これを合成単位データベース作成装置として応用して、複数話者の音声データから、一人の話者の声質をもつ合成単位データから成る合成単位データベースを作成することも可能である。このデータベースを用いる場合には、音声合成装置としては、実施例の構成において、合成単位データ情報テーブルから話者IDを省略することができ、また、声質変換部及び個人コードブック格納部も省略することができる。但し、合成単位データ選択の際には、データベース作成の際に計算した声質(話者)の一致度(距離)を考慮した方が、データベース作成時の声質変換によって大きく品質の低下してしまった合成単位データを選択してしまうおそれがあるため、この距離を合成単位データ情報テーブルに記述において評価関数に導入するようにした方がよい。

【0025】

【発明の効果】本発明によれば、複数の人間から収録した音声データを合成単位データとして蓄積しておき、合成に使用したい合成単位に近い合成単位データを選び出し、これと合成したい声質とがある程度以上異なる場合には、選び出した合成単位データの声質を目標の声質に変換するようにしているので、複数の人間の音声データを利用して所望の声質の合成音声を得られ、従って、一人の人間から大量の音声データを収録する必要がなくなり、収録の労力軽減に効果がある。

10 【0026】また、すでに収録されている複数人の音声データを再利用することができるので、合成単位データベースの作成作業の軽減にも効果がある。

【図面の簡単な説明】

【図1】本発明の声質変換型音声合成装置の一実施例の全体構成を示すブロック図。

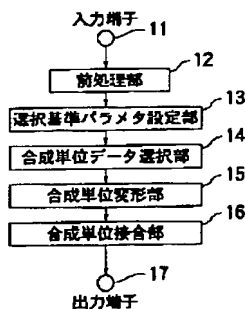
【図2】図1における合成単位データ選択部の詳細構成を示すブロック図。

【図3】図2における合成単位データ情報テーブルの一例を示す図。

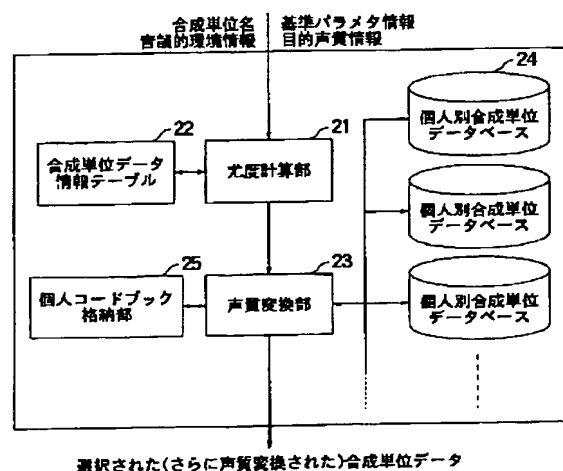
20 【符号の説明】

- 11 入力端子
- 12 前処理部
- 13 選択基準パラメタ設定部
- 14 合成単位データ選択部
- 15 合成単位変形部
- 16 合成単位接合部
- 17 出力端子
- 21 尤度計算部
- 22 合成単位データ情報テーブル
- 23 声質変換部
- 24 個人別合成単位データベース
- 25 個人コードブック格納部

【図1】



【図2】



(5)

特開平7-319495

【図3】

| 合成単位名 | 先行音韻 | 後続音韻 | パワー | ピッチ | 語者 | データID |
|-------|------|------|------|------|-----|----------|
| ... | ... | ... | ... | ... | ... | ... |
| /ka/ | /a/ | /u/ | 24.2 | 10.5 | A01 | 29304156 |
| /ka/ | /a/ | /i/ | 21.2 | 3.5 | A02 | 29304157 |
| ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... |

THIS PAGE BLANK (USPTO)